

CRT Toolkit

A guide to multi-site cluster randomized trials
in the HMO Research Network



Purpose and Description

This guide was developed to help investigators and project coordinators in the HMO Research Network organize and conduct cluster randomized trials (CRTs). It highlights key planning and operational challenges in multi-site CRTs and offers recommendations for how those challenges should be addressed. Content areas include:

- general considerations and rationale
- defining clusters
- issues related to cluster randomization, sample size, informed consent, blinding, and other topics that merit unique consideration in a CRT

Authors

Manel Pladevall, MD, Center for Health Services Research, Henry Ford Health System
Janine Simpkins, Center for Health Services Research, Henry Ford Health System
Allan Donner, PhD, University of Western Ontario, London, Ontario
David R. Nerenz, Center for Health Services Research, Henry Ford Health System

Comments or Questions

If you have questions about HMORN multi-site research not sufficiently addressed here, please refer to your local site administrator.

If you have specific questions or comments about this Toolkit, please contact Ella Thompson at Thompson.e@ghc.org.

Acknowledgments

The development of this guide was funded by the National Institutes of Health, under Contract No. HHSN268200425212C, "Re-Engineering the Clinical Research Enterprise."

TABLE OF CONTENTS

What is a Cluster Randomized Trial (CRT)?	2
Why Cluster Randomized Trials?	3
Figure 1. Summary of reasons for conducting cluster randomized trials	3
General Considerations	4
Rationale for a Cluster Randomized Trial	4
Defining Clusters	5
Defining Clusters in Multi-site CRTs	7
Clarifying Units of Randomization vs. Units of Inference	7
Table 1. Comparison of levels of intervention and levels of evaluation	8
Randomization and Enrollment issues – Minimizing Biases	9
Figure 2. An example of the recommended format for flow diagram of the progress of cluster and individuals through phases of a randomized trial	10
Inclusion-exclusion criteria	11
Sample Size Considerations	12
Efficacy vs. Effectiveness Studies	12
Trial Stratification	13
Blinding	14
Informed Consent	15
Interim Safety and Efficacy Reporting Group Designation	17
References	18

APPENDICES (attached separately)

[Appendix A: Table of cluster characteristics within the HMORN](#)

[Appendix B: Example tracking template](#)

[Appendix C: Example study introduction and consent letters to clusters and individuals](#)

[Appendix D: CRT consent checklist](#)

What is a Cluster Randomized Trial (CRT)?

The idea of intervening at one level involving multiple individual study participants and measuring at another level (e.g., individual plan members) exemplifies the essence of a cluster randomized trial. Cluster randomized trials represent an increasingly popular way to study a variety of clinical and administrative innovations. Organizations like the HMO Research Network represent attractive environments for multi-site CRTs, since they can bring a large number of potentially comparable clusters (e.g., primary care physician practices, clinics) into a planned CRT relatively efficiently.

Let's imagine that you're working with the CEO of a health insurance company and she is wondering whether increasing the pharmacy benefits of patients with diabetes might in the long run be cost saving. You want to conduct a formal study to test that hypothesis. Specifically, you want to randomly assign patients to different benefit levels but you are not sure how to implement this idea. If you randomize individual plan members, you'll need consent from each one, and it's hard to imagine that a plan member would agree to pay more or less for care by a random decision.

However, let's assume that your health insurance company is active in several states or regions. You could then plan a study in which you would implement one benefit level in some states/regions and the other in the remaining states/regions. You would still need consent from someone at the state level with public representation (i.e., the insurance commissioner for a state, or perhaps a plan member representative to a regional advisory board), but you would be in a position to intervene at one level (the state or region) and measure outcomes at another level (individual plan members with diabetes).ⁱ

What about joining forces with other health plans that might be interested in testing the same hypothesis? In this example, because the level of randomization includes large geographic units and you might have not enough units within your own company, you might want to collaborate with other health plans in order to have an adequate test of your hypothesis. You'd now be planning a multi-site cluster randomized trial.

In both of these examples, the planned intervention occurs at a "cluster" level. The states/regions referred to above are examples of what constitutes a cluster - a group of individuals in which a planned study intervention can be implemented. Those clusters become the units that will be randomly assigned to the intervention (enhanced pharmacy benefits) or to the control group (current pharmacy benefits). All of the individuals belonging to that cluster (individual plan members, in this case) receive the same treatment.

In our example, the unit of intervention is the state because that is the natural, or perhaps only, organizational level at which it is possible or practical to manipulate insurance benefit levels. For other types of studies (e.g., studies of physician reminder systems built into electronic medical records, comparative effective trials of medications), a smaller organizational unit would serve well as a cluster. For example, a small group of physicians practicing together (a clinic or a practice) would be a useful definition of a cluster for this type of study.

ⁱ The issue of whether consent from individual participants is required in a study like this is an important and challenging one that will be taken up in detail later in the document.

Why Cluster Randomized Trials?

Cluster randomized trials (CRTs) offer unique opportunities to test interventions that cannot easily be tested in traditional randomized controlled trials.^{1,2} The most common study setting in which cluster randomization is considered is one in which there is clear danger of “crossover” or “contamination” because study participants could interact with other study participants, become aware of the existence of the other experimental conditions, and perhaps change their own behavior as a result.³

Another typical setting for a cluster randomized trial is one in which physicians are an essential part of the intervention. For instance, the effectiveness of a new guideline for physicians might be difficult to evaluate if individual patients are randomized to either an intervention group in which the new guideline will be applied or to a control group that will use the old guideline. Will a physician be able to apply the new guideline only to his/her patients in the intervention group? In many instances, it would be difficult for that doctor to do something different for a group of patients once he/she has learned and been trained on the new protocol or guideline. Therefore, if individual patients were randomized, the physician would be likely to carry on some of the aspects, if not all, of the new guideline to the control patients as well and the control patients would be “contaminated” for purposes of subsequent data analysis. Individual randomization for such a trial may also create obstacles to physician recruitment or even raise ethical issues.

CRTs offer a design choice that tends to minimize, if not totally remove, these difficulties.⁴ Furthermore, interventions that are naturally implemented at the health care “unit” level (e.g., enhancements to its systems, changes in health benefit structures) are best studied among those units rather than among individuals, as they are more reflective of how the intervention would be administered in practice, thus maximizing external validity. Since responses among individuals within those clusters would tend to be correlated, though, (i.e., patients of a specific physician might have similar adherence levels to prescribed medication) some efficiency (statistical power) is lost relative to an individually randomized design.⁵

Box 1: Reasons for carrying out evaluations at cluster level

- Public health and health care programs are generally implemented at organization rather than individual level, so cluster level studies are more appropriate for assessing the effectiveness of such programs.
- It may not be appropriate, or possible in practice, to randomize individuals to intervention groups since all individuals within a general practice or clinic may be treated in the same way.
- “Contamination” may sometimes be minimized through allocation of appropriate organizational clusters to intervention and control groups. For example, individuals in an intervention group might communicate a health promotion message to control individuals in the same cluster. This might be minimized by randomizing whole towns to different interventions.
- Studies in which entire clusters are allocated to groups may sometimes be more cost effective than individual level allocation, if locating and randomizing individuals is relatively costly.

Figure 1. Summary of reasons for conducting cluster randomized trials.⁶

There are unique challenges involved in CRTs.⁷ These include the choice of randomization unit, issues of informed consent and enrollment, and the need to account for clustering effects both in the estimation of sample size and the statistical analysis. All present distinct, but generally manageable, challenges.

The purpose of this document is to highlight the key planning and operational challenges in *multi-site* cluster randomized trials, and offer recommendations for how those challenges should be addressed. Groups like the HMO Research Network (HMORN) are attractive sites for such trials, because they create the opportunity to identify relatively large numbers of similar clusters in an environment that is familiar with collaborative, multi-site studies. Our document should therefore serve as a guide to investigators who are planning such trials, as well as to project coordinators and others who are responsible for implementing them in specific study sites.

General Considerations

Systematic reviews of published CRTs show that methodological errors are common.^{8,9} Therefore, special care is required when using this design to ensure that the right methods are used. Examples of areas that require attention are:

- having a clear rationale for choosing this design
- defining the randomization unit (e.g., physician, clinic, school), and the unit of inference (individual or cluster) to whom the results are to be applied
- documenting the flow of clusters and individuals in the recruitment and implementation phases of the trial
- procedures for informed consent.

Guidelines on how to *report* randomized clinical trials are available. Indeed, many biomedical journals have adopted them and require manuscripts reporting the results of clinical trials to strictly follow them. The Consolidated Standards of Reporting Trials (CONSORT) statement is one of the most well-known guidelines and was first published in 1996.¹⁰ The statement includes a checklist of items that should be included in the trial report. These items are evidence-based whenever possible and are regularly reviewed. In 2004, an extension of these guidelines specific for reporting cluster-randomized trials was published.¹¹ This extension also provides indirect guidance on the relevant issues to consider when designing and implementing CRTs. There are also good reference books that describe in detail how to design and apply sound analytical methods for CRTs.^{12,13,14}

Rationale for a Cluster Randomized Trial

Randomization of intact clusters of patients is generally less efficient than the randomization of individual patients. Therefore, the rationale for choosing cluster randomization should be explicitly stated in the process of planning a trial and in all subsequent written materials. As discussed above, common reasons include administrative convenience, concerns regarding experimental contamination, and ethical

concerns. The rationale for a multi-site study should also be explicitly stated, even though the rationale is usually the issue of adequate numbers of clusters available for randomization.

To determine whether a CRT design is appropriate, it may be useful to consider the following questions:

- Does the thing being studied occur naturally at a “group” or “cluster” level (e.g., a medication guideline that is implemented throughout a group practice), with effects measurable at an individual patient level?
- If individual participants (e.g., patients) were randomized, would it be difficult for others (e.g., physicians) to change behavior according to the arm of the study to which an individual had been assigned?
- Is it likely that individuals in the study (patients, clinicians, other staff) would have occasion to talk among themselves about the study and possibly “cross over” to the other arm (e.g., adopt a diet or exercise plan to which others had been assigned) on the basis of those conversations?
- Is it possible that individuals in the study could intentionally bias results of the study because of their knowledge of the characteristics of different study arms?
- Would it be substantially easier or more efficient to apply the experimental intervention to clusters or units of individual participants rather than to individuals one at a time?

“Yes” answers to one or more of the questions suggest that a CRT design may be appropriate.

Defining Clusters

In health care settings, there are several organizational levels that might be selected as a cluster for purposes of organizing a CRT. These include: individual clinician; clinician “team”; clinic or practice; department; inpatient unit; hospital; defined geographic area (e.g., county or region); employer group; or health plan. Choosing the right level (randomization unit) and considering the implication of this choice on other aspects of the trial design depends on working through a set of criteria that include:

- **Nature of the intervention** - A study of the quality of doctor-patient communication regarding the benefits of cancer screening would typically involve manipulation of behavior at the individual physician or perhaps clinical team level; a study of different co-payment levels for medications would typically involve manipulation of policy at the employer or health plan level.
- **Potential for contamination across study arms** - In a study of electronic reminders to physicians about screening tests for patients, are physicians who practice in the same physical location likely to talk to each

other about the study and their experiences with the electronic reminders? If yes, then the randomization should be done at the clinic or practice level so that all physicians who practice together and talk to each other are in the same study arm. If no, then the study could potentially be done with clustering only at the individual physician level.

- **Number of clusters available for randomization** - Consider a study of changing co-payment levels: if the unit of randomization is the employer, but there are only four large employers served by the health plan, then the number of resulting clusters per study arm is going to be too small and the study will not have sufficient power. To achieve a sufficiently large sample size in a situation like this, it may be necessary to consider smaller units of randomization, even at the risk of some experimental contamination.
- **Identification of individual participants within clusters** - It is essential to have clear relationships between individual study participants and clusters. For example, in a study of cancer screening guidelines and their effects on screening at the individual patient level, randomizing at the individual primary care physician level would not work if patients are not clearly and consistently assigned to PCPs for care. In a situation where patients belong to a team or clinic and are seen by many different providers within the team, then the team or clinic would have to serve as the randomization unit.
- **Characteristics of clusters** - If the clusters are not expected to be similar regarding important outcome predictors, then *stratified randomization*, which creates subgroup of clusters (strata) that are relatively homogeneous, should be considered. Clusters must then be defined in a way that information on potential predictors is available and uniformly defined from cluster to cluster. Commonly adopted stratification factors include cluster size and geographic location. In trials randomizing medical units (either clinics or individual clinicians), years of experience of physicians, solo vs. group practice, and physician gender could also be considered.
- **Number of available individual observations per cluster** - Clusters of relatively large size are less likely to be prone to problems of experimental contamination than are clusters of small size. However, this advantage comes at the cost of statistical power, since randomization of a large number of small clusters tends to be more efficient than randomization of a small number of large clusters. For example in an early trial reported by Ferebee¹⁵, the unit of randomization was selected to be a group of combined hospital wards rather than individual wards, thus reducing the administrative complexity of the study and possibly reducing the risk of experimental contamination.

Defining Clusters in Multi-site CRTs

Defining what constitutes a cluster in health care settings is especially challenging when designing *multi-site* cluster-randomized trials. For instance, primary care physicians might see children in one health plan but not in another. A practice might be defined as a small or very large group of doctors in different organizations. Multiple practices might work in the same building in one health plan or in separate building in another health plan. Some hospitals might have residents and others not.

To the extent that variables like primary care case mix, number of physicians practicing together, and presence of residents in hospital can affect a study intervention, the definition of cluster has to be adjusted to the features of participating organizations (study sites) to maximize comparability of clusters from site to site. For example, a study of physician incentive payments for adherence to screening guidelines might be designed so that only physicians seeing both children and adults are eligible to be defined as a “cluster,” even though many primary care physicians might be excluded by that criterion at some study sites. It would also be appropriate to only include clusters in which groups of physicians had the same general class of financial incentives (e.g., individual vs. group incentives).

All the aforementioned cluster definitions are valid as long as they are defined consistently across sites. If the clusters are different conceptually across sites, then the unit of randomization is not clear, and this will affect the comparability of the groups and the study validity. Therefore, it will be very important at the design phase to ensure that clusters are consistently defined across the different sites.

The table included as [Appendix A](#) identifies several characteristics of potential units of randomization in the HMO Research Network. More information may be needed for specific studies, but the table identifies a number of parameters at the physician, practice, region, or whole plan level that may or may not be comparable across potential study sites.

Clarifying Units of Randomization vs. Units of Inference

Frequently, the unit of inference in a cluster randomization trial is defined at the individual level, while the unit of randomization is at the cluster level. This discrepancy between the unit of randomization and the unit of inference creates the need to adjust for clustering effects, as discussed in the text by Donner and Klar.¹³ This adjustment needs to be made during the study planning process as well as later on, during statistical analysis. For example, in evaluating the effect of a new procedure for treating obesity, medical practices rather than individual patients might be selected as the unit of randomization. This will allow a physician to treat all of his/her patients uniformly while removing the randomization procedure as an impediment to the daily practice workflow. The main purpose of the intervention in such a trial, though, may be to affect patient outcomes such as weight loss at the *individual* level. If so, a difference in weight loss at the individual level would be the main hypothesis to be tested in the study, and individual weight loss would be the primary dependent variable in most analyses. The influence of individual-level covariates such as age and gender on the intervention effect may also be of interest.

In other trials though, particularly those developed from a policy perspective, the units of randomization and of inference may be the same. For example, Diwan et al¹⁶ evaluated a policy of “group detailing” on the prescribing of lipid-lowering drugs in a trial randomizing community health centers. A primary endpoint in this study was the number of appropriately administered prescriptions per month, with the health center serving as the unit of analysis. In trials such as these, where the observed results on any one subject are not of direct interest, but the focus is rather on the implementation of guidelines, the issues that arise in the design and analysis are essentially the same as those that which arise in individually randomized trials. Nevertheless, since clusters rather than individuals are the unit of randomization in such trials, the terminology “cluster randomized trial” remains appropriate.

Level of evaluation	Level of intervention	
	Individual	Area or organization
Individual	Clinical trial—for example, does treating multiple sclerosis patients with interferon beta reduce their morbidity from the condition?	Area or organization-based evaluation—for example, does providing GPs with guidelines on diabetes management improve blood glucose control in their patients? Does providing a “baby friendly” environment in hospital increase mothers’ success at breast feeding?
Area or organization		Area or organization based evaluation—for example, do smoking control policies increase the proportion of smoke free workplaces? Do fundholding general practices develop better practice facilities than non-fundholders?

Table 1. Comparison of levels of intervention and levels of evaluation⁶

Once the units of randomization and units of analysis have been identified, a key design decision in cluster randomized trials is whether the intervention is to be *delivered* to the cluster as a whole, as in policy or guideline trials, or directly administered to individual participants. Although the CRT design typically implies intervention at a cluster or unit level, it is possible to have a CRT in which the study intervention is actually delivered to individual participants one at a time. The cluster design and approach to analysis are still appropriate when all individuals in a cluster receive the same treatment, even if that treatment is actually given one participant at a time.

The nature of the intervention typically dictates how it is delivered. Interventions that inherently function at a “cluster” level—e.g., use of a clinical guideline, enhancement of a feature of an electronic medical record, change in a health insurance benefit—would naturally be delivered at the cluster level in the context of a study. In fact, there are many instances when there may be no other effective way to do the study.

In other words, one may think of the intervention as being indivisible at the cluster level. On the other hand, interventions that are naturally directed to individual patients within a physician cluster (e.g., health education messages), or to individual physicians within a clinic cluster (e.g., academic detailing, financial incentives for quality improvement) could be delivered to study participants on a one by one basis.

Randomization and Enrollment issues – Minimizing Biases

To strengthen internal validity, it is desirable (although not always possible) to identify all eligible clusters and individual participants and obtain informed consent prior to randomization. Any biases produced by cluster-level or individual-level decisions not to participate would then be spread equally across study arms, assuring an unbiased test of the study hypothesis.¹⁷ (Issues of external validity would still remain, though, if the characteristics of clusters or individual agreeing to participate were markedly different from the population from which they were drawn.) Meeting this condition would also ease the task of generating a clear diagram showing the flow of clusters and participants through a trial (see [Figure 2](#) as an example).

In some studies, it may be possible to identify in advance all individuals in a cluster that will be included in the proposed trial. This might occur, for example, in a trial of enhancements to an electronic medical record system in which consent from individual patients is not required (i.e., only administrative data will be used), all individual patients in each cluster can be identified, and all will be included in the study. In this case, randomization of a sufficient number of clusters will assure that both cluster-level and individual-level characteristics will be well-balanced between groups. In other words, there will be no risk of selection bias at either of these levels.

However, in some trials, patients must be enrolled over a predetermined accrual period as they appear for treatment (i.e., a study of guidelines for the assessment of patients with newly diagnosed heart failure). In this case, consent cannot be obtained from individual patients for randomization, since randomization at the cluster level has already taken place. Even in studies where all individuals are identified and recruited at the same time, a design that allows individuals or clusters to refuse to participate after randomization has the potential for bias if individuals or clusters agree at different rates in different study arms. From a methodological perspective, internal validity considerations require that the patients are recruited in similar fashion in all study arms, as further discussed below. When eligible subjects can only be identified after their cluster has been randomized, balance with respect to individual-level characteristics can only be guaranteed if subjects entering the cluster over time are sampled at random, or if it is possible to enroll the entire population of individuals.¹⁸

Unfortunately, these are often not realistic options, and patient recruitment is therefore done somewhat opportunistically, thus creating a risk of treatment-related bias at the final evaluation stage. For example, project staff in a medical practice that have been randomized to an exciting or interesting new intervention may be more enthusiastic and motivated in their efforts to recruit individual patients than are

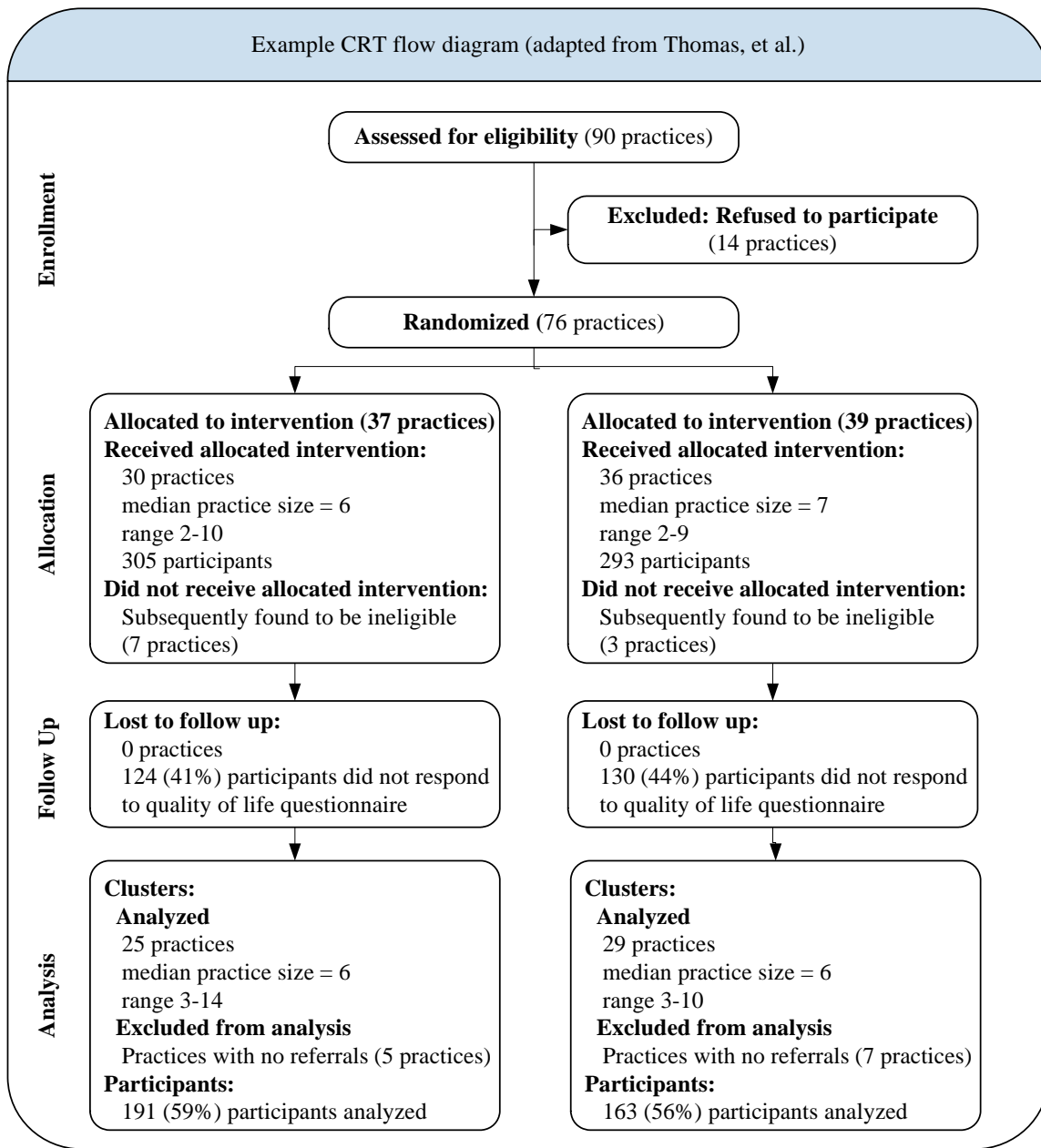


Figure 2. An example of the recommended format for flow diagram of the progress of cluster and individuals through phases of a randomized trial¹⁰

staff who are assigned to the “usual care” control group. Torgerson¹⁹ has shown that such differential recruiting efforts can result in a systematic and possibly substantial imbalance between groups with respect to both the number of subjects and their characteristics. In this case much of the benefit of the original randomization may be lost, since in this case it is only the cluster-level characteristics that will not be subject to selection bias. In the context of trials randomizing medical practices, where it is routine to identify eligible patients after their practice has been randomized, this phenomenon may be referred to as ‘detect and treat bias.’ If eligible patients and clusters are identified before randomization, and some level of data is available from automated data sources, one can at least evaluate the risk

and direction of potential biases. Also, if outcomes data are also available from automated data sources, intention to treat analyses can be performed in all eligible patients within clusters in a similar way as done in regular clinical trials. [Appendix B](#) provides an example of a spreadsheet-based tracking system that could serve as a helpful template. The example was used to track and report on eligible patients, recruited patients, patients actually participating in a trial, and patients completing the trial in a cluster randomized design.

If eligible subjects in a cluster cannot be identified before randomization, it may also be possible to minimize potential unknown biases by ensuring that subject recruitment is done by someone who is blind to group allocation or, at a minimum, someone who is independent of the study. For example, a project coordinator or research nurse based in a research department could handle patient recruitment in a CRT with physician practices as the unit of randomization, preferably in circumstances where he or she was unaware of the treatment arm assignment of any practice from which patients were being recruited.

If it is impossible to obtain informed consent prior to randomization, the persons or process responsible for obtaining consent should keep a log of how many individuals and clusters are eligible, how many are invited to participate, and how many refuse and the reasons. The tracking template provided in [Appendix B](#) is one example of how such a log can be maintained. However, this might not be easy to implement in some cases, for example, when randomization occurs at the physician level and individual doctors identify eligible participants. The ability to detect bias in recruitment in situations like this is enhanced when inclusion and exclusion criteria for individual participants/patients are based on information available in electronic medical records systems or encounter data bases where they can be independently verified. Site Project Coordinators can then check lists of actual study participants identified by physicians against lists of potential participants identified through electronic data systems.

Inclusion-exclusion criteria

Two descriptions of inclusion- exclusion criteria are required, one at the cluster level (e.g., cluster size, geographic area) and the other at the individual level (e.g., age, gender). As in any clinical trial, the inclusion criteria at the cluster level will determine the external validity of the study. For instance, for a comparison of an electronic prescribing system with extra features compared to a conventional system, one will need an inclusion criterion that identifies physicians who are capable of actually using the electronic prescribing system as the only physicians eligible for the study. However, the results will then be applicable only to those physicians who meet this eligibility criterion, and that might be a small minority of all physicians.

Inclusion/exclusion criteria at the individual level are generally similar to those used in individual-level clinical trials. The only obvious area of difference is a set of criteria that define the relationship of the individual to the cluster. In a CRT of a revised asthma guideline with the physician practice as the unit of randomization, for example, the eligibility criteria for individual patients would have to include one or

more criteria about the relationship between the patient and the physician. Presence of an ongoing continuing care relationship, length of time in the practice, previous occurrence of a number of asthma-related visits, or a formal PCP relationship in the context of managed care are all examples of these sorts of criteria. The specific criteria used in individual studies will vary; it is helpful to identify what relationships between patients and the unit of randomization must be in place in order to produce a fair test of the study hypothesis.

Sample Size Considerations

Ignoring clustering effects in the estimation of sample size is likely to lead to an underpowered trial (elevated type 2 error).²⁰ Thus, findings in trials that have ignored clustering effects at the design stage may be declared “nonsignificant” or “no difference,” when in fact they would be more accurately characterized as inconclusive. The number of clusters, the average number of subjects expected per cluster, the likely size of the anticipated clustering effect, and the anticipated loss to follow-up rate (at both the cluster and individual level), should all be included in the assessment of trial size. The usual method for accounting for the clustering effects in this assessment is to multiply standard sample size formulas by a factor that quantifies this effect (often known as the “design effect”).²¹ Information on the likely size of the design effect could be obtained from reports of previous trials using similar randomization units and primary endpoints. If no such previous studies are available, then a conservative approach is recommended, in which relatively large values for the intraclass correlation coefficient and design effect are used to calculate sample size. There are a number of excellent background articles available on this point.^{22,23,24}

The power of a cluster randomized trial depends more on the number of clusters than on their size. The number of individual participants (e.g., patients) per cluster is relatively less important. A subsampling strategy for individuals within clusters may be adopted when very large units such as hospitals are randomized and costs of including large numbers of individuals in the study are relatively high (for example, when the individual is the direct recipient of the intervention or when the primary data of interest come from individual interviews or surveys or medical record reviews rather than from existing administrative data systems). In these situations, a pre-determined number of subjects can be randomly chosen in each cluster to participate in the study. Careful specification of the individual-level eligibility criteria is then required, balancing concerns for external validity with the need for internal validity. A tracking data base like that provided in Appendix B is a useful way to track the sequence of individuals eligible, sampled, contacted, agreeing to participate, and finally completing the study.

Efficacy vs. Effectiveness Studies

Cluster randomized trials, like individually randomized trials, may be broadly characterized as either “efficacy or “effectiveness” studies, depending on whether the interventions are tested under “ideal” or “real world” conditions, respectively. In the former case, both clusters and subjects are selected according to restrictive criteria, with the intervention administered in similar fashion to all subjects. The eligibility criteria in effectiveness trials may be much looser, reflecting the desire to generalize the results to a relatively heterogeneous population, with the intervention possibly “tailored” to either

the individual clusters or individual subjects.²⁵ The former approach maximizes internal validity (strength of causal inferences) at the expense of external validity (ability to generalize findings to routine medical practice settings); the latter approach maximizes external vs. internal validity.

In the context of CRTs, there are no firm rules or guidelines about when efficacy vs. effectiveness studies are preferred. Both types of studies can make a valuable contribution to a research area, and it may be possible to design studies that represent a compromise or hybrid sort of design that balances considerations of internal vs. external validity. In most research areas, efficacy studies that address questions of whether something can have an effect under ideal, controlled circumstances precede effectiveness studies that address questions of whether something that works in ideal settings can also work in less well-controlled settings.

Trial Stratification

It is possible to design and conduct CRTs in which clusters are randomized from the entire available universe of clusters. In many instances, though, characteristics of clusters that can be known in advance may influence study outcomes, and randomization of relatively small numbers of clusters may not provide adequate assurance of balance or those characteristics across study arms. For example, in a CRT that involves randomization of 20 primary care practices to treatment vs. control arms, an investigator may be concerned that the practices have widely varying payor mixes and SES distributions in their patient populations. It is possible that simple randomization may assign most of the “rich” practices to the experimental arm and most of the “poor” practices to the control arm. In a situation like this, some protection is provided by a stratified or matching design in which randomization takes place within pre-defined strata or within pre-defined cluster pairs.

In a cluster randomization trial, strata are necessarily defined at the cluster level. Commonly adopted stratification factors include cluster size, geographic area, and baseline rates of important process or outcome variables (for example, current rates of adherence to clinical process guidelines in a study designed to enhance adherence to guidelines). The latter (process and outcome variables relevant to the study intervention) are likely to be the most effective choice from a precision standpoint provided the appropriate data are available. However, stratification by cluster size (e.g., number of physicians in a practice) is also very desirable since it may be a surrogate for within-cluster dynamics that are predictive of the study endpoints. For example, smaller clusters may lead to increased interaction among cluster members, which in turn may lead to better compliance with an intervention whose effects depend on visibility and awareness among cluster members. Stratification by cluster size also assures reasonable balance in the number of individuals assigned to each treatment group, which impacts efficiency.

The table of cluster characteristics within the HMORN ([Appendix A](#)) should be a useful tool when considering a stratified study design in this environment. Many potential stratifying variables for different types of cluster definitions are included in the table.

A special case of the stratified design arises when each stratum contains only two clusters, one assigned to the experimental intervention and one to the control. Such a

pair-matched design provides tighter control of potential confounding factors, although it may not always be possible to obtain “close matches.” Pair-matched designs also raise unique analytical issues, as discussed by Donner and Klar¹³ (Chapter 3).

Blinding

In individual clinical trials, double-blinding (neither the physician nor the patient knows which experimental treatment the patient is receiving) is the gold standard. Double-blinding eliminates many of the opportunities for biased assessment of outcome that could occur when either the clinician or patient desires a particular study outcome and there is an element of subjective judgment in rating outcomes.

Bias in the identification and recruitment of individual study participants within clusters is one potential problem that is influenced by procedures for blinding. Jordhay et.al.¹⁷ describe how serious problems of selection bias arose in a cluster randomized trial of palliative care that did not involve blinded group allocation. Since it is usually more difficult to double blind non-pharmacological interventions, it is important to keep investigators unaware of the randomization scheme (allocation concealment) and to ensure that there is no room for manipulation by investigators of the randomization scheme and hence assignment of a patient or cluster to the “preferred” intervention. To minimize this risk, randomization could be performed by a third party that has no direct relationship with the investigators. In a multi-site setting, the coordinating center should be responsible for that process. If the study involves only one site, external parties such as an “independent statistician” may be employed.

If a representative of a cluster (e.g., a project manager in a clinic) is responsible for identifying individual study participants (e.g., individual patients), a system for temporarily blinding the project managers to study condition—until the step of identifying individual patients has occurred—can reduce possible bias in recruitment. However, if it is an investigator or project staff member outside the cluster who has responsibility for identifying individual trial participants, he/she should be blinded to group membership (i.e., cluster assignment) until all potential subjects are judged as eligible.

In cluster randomized trials, there is frequently an opportunity to not only blind individual participants in a cluster about their treatment assignment, but to avoid description of other study arms entirely. Since all individual in the cluster receive the same treatment, they may not have to be aware that another study condition exists or what its characteristics are. Is this ever ethical?

When the intervention is of a very general nature, as for example, in the case of a health education message, there may be little ethical objection to subjects not being aware of specific elements. For example, in the hypertension screening trial reported by Bass et al.,²⁶ the patients in the experimental and control practices were not aware of the specific nature of the investigation, although they knew that their physician was engaged in a research project. Although such lack of awareness may be viewed from a methodological perspective as design strength, it is clear that ethical concerns must always take priority and be considered separately for each proposed trial.

Informed Consent

The need for informed consent and appropriate procedures for its implementation have been controversial in cluster-randomized trials.^{27,28} These issues are further complicated by the need to consider two or more levels of consent, one or more at the cluster level and one at the individual level. Edwards et al.²⁹ distinguished between individual-cluster trials in which individual members of a cluster can be the targets of intervention, and cluster-cluster trials in which the intervention must be targeted at the cluster as a whole. They point out that, at least in principle, consent could be obtained from individuals in a individual-cluster trial, but that consent from individual cluster members is not possible in cluster-cluster trials. In both cases, they suggest that a “guardian” (e.g., the medical director of a clinic or health plan) is an appropriate person to authorize the use of a cluster in a research project, as long as that person is able to balance risk and benefit for the cluster and make a decision that has the cluster’s best interests as the primary criteria. Eldridge et al extend this point and recommend that lay individuals representing potential trial participants (e.g., lay member of the health plan’s Board, or representatives of a patient advisory panel) be involved in the decision to involve clusters in a trial.³⁰

Provided an appropriate “guardian,” such as a physician or hospital director, can be identified, consent at the cluster level can be relatively straightforward. For example, all eligible physicians representing one or more potential study clusters could be sent a letter of introduction to the study from the site principal investigator and co-signed by the site medical director. This letter could describe the study in general terms and explain what study participation involves as well as the voluntary nature of participation (see example letter in [Appendix C](#)). Those wishing to participate can be asked to notify the principal investigator via a return fax, letter, or telephone call. In some circumstances, where an additional contact would not be viewed as coercive, those who do not notify the principal investigator can be called by the principal investigator to solicit their interest. All physicians who ultimately agree to participate can be sent an informed consent form for review and signature.

When no clear decision-maker is evident, selected representatives from the health care or community leadership could act as surrogates in providing agreement for random assignment. However, it is not clear whether the agreement of such surrogates is sufficient.²⁷

When permission from decision makers associated with each cluster is needed for *assigning interventions*, some indication should be provided in all subsequent descriptions of the study as to who these decision makers were, how they were identified, and how permission was or was not granted for clusters in specific study arms. For example, if regional medical directors gave permission in 100% of instances for clinics to participate in the experimental condition of a CRT but permission was only granted in 50% of instances for clinics to participate in the control condition, concerns about potential biases in the study would be natural. Some information about the consent procedure administered to individual study participants should also be provided. In particular, it would be helpful to know what opportunities, if any, existed for individual cluster members to avoid participation in the intervention.

For those trials in which patients must be approached post-randomization, it is clear that consent can only constitute agreement to be studied under a treatment regimen that has already been assigned. In this sense, the procedures adopted bear close resemblance to those that characterize the design proposed by Zelen³¹ as a means of enhancing recruitment in individually randomized trials. For example, in Zelen's "single-consent" design, consent is sought only from patients who have already been randomized to the experimental group, while no consent at all is obtained from control group patients who receive standard therapy. Although Zelen's design has proved controversial on both methodological and ethical grounds, it is clear that a very similar approach to informed consent has long been recognized as acceptable in trials that randomize clusters.

The following suggestions were proposed in the 1991 International Guidelines for Ethical Review of Epidemiological Studies, put forward by the Council for International Organizations of Medical Sciences³²:

When it is not possible to request informed consent from every individual to be studied, the agreement of a representative of a community or group may be sought, but the representative should be chosen according to the nature, traditions and political philosophy of the community or group. Approval given by a community representative should be consistent with general ethical principles. When investigators work with communities, they will consider communal rights and protection as they would individual rights and protection. For communities in which collective decision-making is customary, communal leaders can express the collective will. However, the refusal of individuals to participate in a study has to be respected: a leader may express agreement on behalf of a community, but an individual's refusal of personal participation is binding.

The previous sentence notwithstanding, there might be instances in which the risk for patients is extremely low and the intervention could be studied without a need for individual consent. For instance, a formal study may be done of a new electronic prescribing system what includes an add-on interface that provides safety alerts and adherence measurements. In the proposed study, physicians will be randomly assigned to either the new interface or the conventional electronic prescribing system.

It may be acceptable to argue that similar systems are already common in other health systems, that the features being studied do not involve any conceivable significant risk to patients, that patients' permission is not required to change features of an electronic medical record system outside the context of research, and that the primary outcome variables (adherence, or occurrence of adverse events) are part of the medical record data routinely available to clinicians to track patients' clinical course. An IRB could consequently decide that individual patient consent is not required.³³ Moreover, study sites in environments like the HMORN may already have a system for informing plan members or patients at the time of enrollment that studies of "administrative improvements" may be going on routinely and that participation in such studies without specific informed consent is a condition of belonging to, and receiving medical care in, the organization.³⁴

For examples of introduction and consent letters for both individuals and clusters, please see [Appendix C](#). For a checklist designed to serve as a guide for investigators in determining the type of consent likely to be required for different types of CRTs, please

see [Appendix D](#). For a more detailed review of issues related to IRB review and need for written informed consent in various types of study design, see Lynn et al.³⁵ and Nerenz et al.³⁶

Interim Safety and Efficacy Reporting Group Designation

The need to perform interim analyses of efficacy in individually randomized clinical trials is largely motivated by ethical concerns. Thus data-dependent stopping rules, such as that developed by O'Brien and Fleming,³⁷ may be used to recommend early termination of a study in the event that one of the interventions shows evidence of superior efficacy before patient follow-up has been totally completed. However, it is notable that while many cluster randomization trials are designed to follow up medically serious endpoints, formal stopping rules have usually not been incorporated at the design stage. This is at least partly because the theory underlying these plans have invariably assumed individual randomization. However, Zou et al.³⁸ have recently shown that the most frequently used stopping rules for individually randomized trials may be adapted to most cluster randomization trials in routine fashion.

There is no reason to presume that the issue of interim safety analysis and/or formal stopping rules is not relevant to CRTs. Although many CRTs may be organized to test relatively safe or benign quality improvement interventions, there is always the possibility that a significant difference in favor of one study arm vs. another may emerge during the study that would justify stopping the study and assigning all clusters to the "favored" intervention, or on the other hand, there may be safety issues (patient or clinician) that arise even in what appear to be benign or harmless interventions. For example, a CRT designed to test the effect of an enhancement to a medical record system may create an unexpected time and effort burden on participating physicians that could have adverse effects on clinicians and patients beyond simply willingness to continue in the study. Investigators should give careful consideration to the idea of creating an interim safety and efficacy monitoring committee, with formal rules determined in advance to stop the study for reasons of either unanticipated adverse events or clear benefit of one treatment condition vs. the others. For more discussion on this issue, see Zou et al.³⁸

References

- ¹ Fayers PM, Jordhoy MS, Kaasa S. Cluster-randomized trials. *Palliat Med.* 2002;16:69-70.
- ² Donner A, Brown KS, Brasher P. A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979-1989. *Int J Epidemiol.* 1990;19:795-800.
- ³ Borm, GF, Melis, RJ, Teerenstra, S, & Peer, PG. Pseudo cluster randomization: a treatment allocation method to minimize contamination and selection bias. *Stat Med.* 2005 Dec 15;24(23):3535-47.
- ⁴ Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Public Health.* 2004;94(3):416-422.
- ⁵ Xies, T, Waksman, J. Design and sample size estimation in clinical trials with clustered survival times as the primary endpoint. *Statistics in Medicine*, 2003, 22, 2835-2846.
- ⁶ Ukoumunne OC, Gulliford MC, Chinn S, et al. Methods in health service research: Evaluation of health interventions at area and organisation level. *BMJ*, 1999, 319, 376-379.
- ⁷ Lee, KJ & Thompson, SG. Clustering by health professional in individually randomized trials. *BMJ*, 2005, 330, 142-144.
- ⁸ Murray, DM, Varnell, SP, & Blitstein, JL. Design and analysis of group-randomized trials: A review of recent methodological developments. *Am J Pub Health*, 2004, 94(3), 423-432.
- ⁹ Varnell, SP, Murray, DM, Janega, JG, & Blitstein, JL. Design and analysis of group-randomized trials: A review of recent practices. *Am J Pub Health*, 2004, 94, 393.
- ¹⁰ Moher, D, Schulz, KF, Altman, DG, & Lepage, L. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*, 2001, 357, 1191-1194.
- ¹¹ Campbell, MK, Elbourne, DR, Altman, DG. CONSORT statement: Extension to cluster randomised trials. *BMJ*, 2004, 328, 702-708.
- ¹² Medical Research Council. *Cluster Randomized Trials: Methodological and Ethical Considerations*. MRC Clinical Trials Series. London: Medical Research Council, 2002.
- ¹³ Donner, A & Klar, N. *Design and Analysis of Cluster Randomization Trials in Health Research*. London: Arnold, 2000.
- ¹⁴ Murray, DM. *Design and Analysis of Group-Randomized Trials*. New York: Oxford University Press, 1998.
- ¹⁵ Ferebee SH, Mount FW, Murray FJ and Livesay VT. A controlled trial of isoniazid prophylaxis in mental institutions. *American Review of Respiratory Disease*, 1963, 88:161-175.
- ¹⁶ Diwan VK, Wahlstrom R, Tomson G, et al. Effects of "group detailing" on the prescribing of lipid-lowering drugs: a randomized controlled trial in Swedish Primary Care. *J Clin Epidemiol* 1995;48:705-711.
- ¹⁷ Puffer, S, Torgerson, DJ, & Watson, J. Cluster randomized trials. *J Eval in Clin Pract*, 2005, 11, 479-483.
- ¹⁸ Puffer, S, Torgerson, DJ, & Watson, J. Evidence for risk of bias in cluster randomized trials: Review of recent trials published in three general medical journals. *BMJ*, 2003, 327, 785-789.
- ¹⁹ Torgerson, DJ. Contamination in trials: Is cluster randomization the answer? *BMJ*, 2001, 322, 355-357.
- ²⁰ Hsieh, FY. Sample size formulae for intervention studies with the cluster as unit of randomization. *Stat Med*, 1988, 8, 1195-1201.

- ²¹ Eldridge, S, Kerry, S, & Ashby, D. Sample size for cluster randomized trials: Effect of coefficient of variation of cluster size and analysis method. *Int J Epid*, 2006, 35, 1292-1300.
- ²² Donner, A. Sample size requirements for interval estimation of the intraclass kappa statistic. *Commun Stat Simulation Computation*, 1999, 28, 415-428.
- ²³ Smeeth, L, Suiwoon, NE. Intraclass correlation coefficients for cluster randomized trials in primary care: Data from the MRC Trial for the Assessment and Management of Older People in the Community. *Control Clin Trials*, 2002, 23, 409-421.
- ²⁴ Campbell, MJ. Cluster randomized trials in general (family) practice research. *Stat Methods Med Res*, 2000, 9(2), 81-94.
- ²⁵ Glasgow, RE, Magid, DJ, Beck, A et al. Practical clinical trials for translating research to practice: Design and measurement recommendations. *Med Care*, 2005, 43, 551-557.
- ²⁶ Bass, MJ, McWhinney, IR, Donner, A. Do family physicians need medical assistants to detect and manage hypertension? *CMAJ*, 1986, 134, 1247-1255.
- ²⁷ Winkens, RA, Knottnerus, JA, Kester, AD, et al. Fitting a routine health-care activity into a randomized trial: An experiment possible without informed consent? *J Clin Epidemiol* 1997, 50, 435-439.
- ²⁸ Hutton, JL. Are distinctive ethical principles required for cluster randomized trials? *Stat Med* 2001, 20, 473-488.
- ²⁹ Edwards, SJL, Braunholtz, DA, Lilford, RJ, & Stevens, AJ., Ethical issues in the design and conduct of cluster randomized controlled trials. *BMJ*, 1999, 318, 1407-1409.
- ³⁰ Eldridge, SM, Ashby, D, & Feder, GS. Informed consent to participation in cluster randomized trials: An empirical exploration of trials in primary care. *Clinical Trials*, 2005, 2, 91-98.
- ³¹ Zelen, M. Randomized consent designs for clinical trials: An update. *Stat Med*, 1990, 9, 645-656.
- ³² Brody, BA. *The Ethics of Biomedical Research: An International Perspective*. New York, NY: Oxford University Press, 1998. (pp 225-226.)
- ³³ Mazor, KM, Sabin, JE, Boudreau, D, et al. Cluster randomized trials: Opportunities and barriers identified by leaders of eight health plans. *Med Care*, 2007, 45(10, Suppl 2), S29-S37.
- ³⁴ Goldberg, HI, Neighbor, WE, Hirsch, IB, et al. Evidence-based management: using serial firm trials to improve diabetes care quality. *Jt Comm J Qual Improv*. 2002 Apr;28(4):155-66.
- ³⁵ Lynn, J, Baily, MA, Bottrell, M, et al. The ethics of using quality improvement methods in health care. *Annals of Internal Medicine*, 2007, 146(9), 666-673.
- ³⁶ Nerenz, DR, Stolz, P. K., and Jordan, J.. Quality improvement and the need for IRB review. *Quality Management in Health Care*, 12, 159-170.
- ³⁷ O'Brien, PC & Fleming, TR. A multiple testing procedure for clinical trials. *Biometrics*, 1979, 34, 549-556.
- ³⁸ Zou, G, Donner, A, & Kar, N. Group sequential methods for cluster randomization trials with binary outcomes. *Clin Trials*, 2005, 2, 479-485.